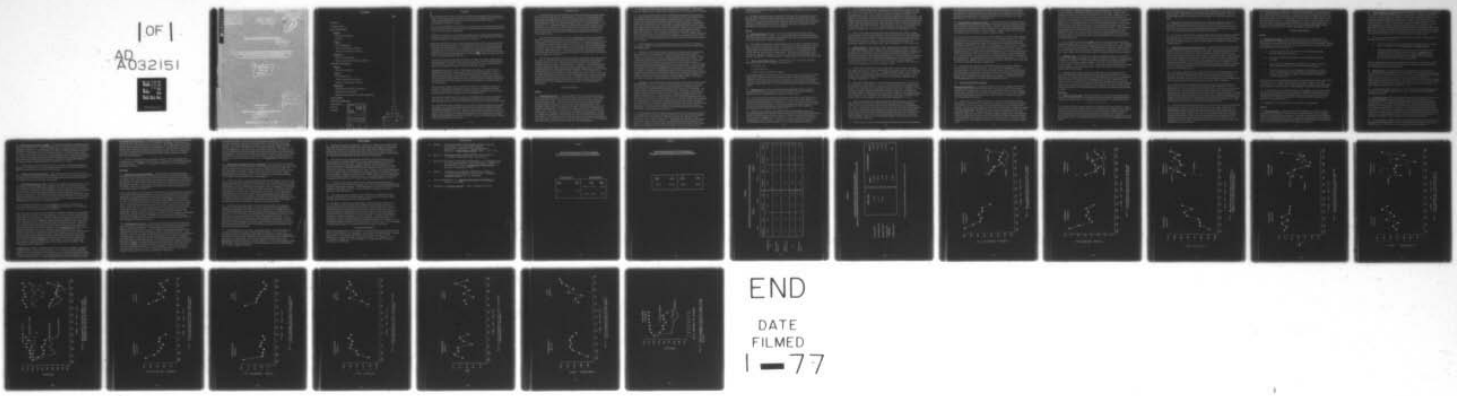
AD-A032 151   ROYAL NAVAL PERSONNEL RESEARCH COMMITTEE LONDON (ENGLAND)   F/G 6/19
SIGNAL DETECTION EFFICIENCY IN THE MORNING WATCH. EFFECTS OF PR--ETC(U)
MAR 74   W P COLQUHOUN, P HAMILTON, R S EDWARDS
UNCLASSIFIED   OES-10/74   DRIC-BR-51850   NL

1 OF 1
AD
A032151



END
DATE
FILMED
1 — 77

CONTENTS

	Page
SUMMARY	
INTRODUCTION	1
MAIN EXPERIMENT	1
Method	1
- Design and procedure	1
- The task	2
Results	3
- Body temperature	3
- "Basic" performance measures	3
- "Derived" performance measures	5
Discussion	6
- Effects of staying awake	6
- Effects of time of watch and task duration	7
CONTROL EXPERIMENT	8
Method	8
- Design and procedure	8
Results	8
- Body temperature	8
- "Basic" performance measures	9
- "Derived" performance measures	10
Discussion	11
- Within-watch variations in efficiency	11
- Performance over the day	11
- Relevance to the results of the main experiment	12
CONCLUSIONS	13
ACKNOWLEDGEMENTS	13
REFERENCES	14
TABLES	15 - 18
FIGURES	19 - 30

SUMMARY

An experiment was carried out to determine whether staying awake before the morning (0400 - 0800) watch (a custom observed in submariners on prolonged patrols) is likely to exert a detrimental effect on operations such as sonar monitoring carried out continuously throughout the watch.

The performance efficiency of 12 ratings on a signal detection task was monitored during the morning watch on four separate occasions; two without prior sleep and two after four hours' sleep immediately beforehand.

Although in the early part of the watch there appeared to be certain slight advantages in having remained awake, during the second half of the watch the signal detection rate was consistently higher when sleep had been taken beforehand.

Even when sleep had been taken, general performance levels in the morning watch were markedly different from those in a watch held from 2000 to 2400 the previous evening. Detection rate was subtantially lower throughout the watch, and the degradation during the second half of the watch at times exceeded 50 per cent when compared to "fresh" performance at the start of the previous evening watch. Accompanying changes in other aspects of performance suggested that these strikingly large decrements were caused by failure to maintain attention to the task.

In order to determine to what extent this failure resulted from fatigue arising from the continuous nature of the task and the allowance of only four hours rest between watches, the performance of a further 11 ratings was monitored on the same task during "day" hours, on four separate occasions, following a full night's sleep.

Although the same 4-ON, 4-OFF, 4-ON schedule was adhered to in this "control" experiment, performance during the dog watch (1600-2000) was found to be in no way different from performance in the preceding forenoon watch (0800-1200). However, comparison with results obtained from earlier watchkeeping studies indicated that this absence of between-watch change concealed an underlying effect of fatigue; estimates of the magnitude of this fatigue effect were then made and applied to the results from the main experiment.

It was shown that fatigue interacted with diurnal rhythm to considerably enhance the changes in performance resulting from the latter alone; the combination of effects resulted in a decrement in detection rate in the morning watch which was $2\frac{1}{2}$ to 3 times greater than that found when fatigue was not a confounding factor.

The results also demonstrated, once again, the substantial nature of the decrements in efficiency that occur in any case during continuous performance of a detection task, even when this task is carried out in the absence of any stress arising from loss of sleep.

It is concluded that task-interruption or job-rotation can mitigate both this within-watch decrement, and also the detrimental effects of diurnal rhythm on efficiency during the morning watch; performance in this watch can also be improved by obtaining at least 8 hours rest beforehand. However, completely stable levels of efficiency over the night hours can only be achieved by permanent assignment of certain personnel to standing night watches throughout the duration of a patrol, in order to allow physiological adaption to such unusual hours of work to occur.

INTRODUCTION

1. Previous studies of operational efficiency under different watch-keeping systems (Ref 1) have shown that the level of performance at "mental" (simulated sonar and communications) tasks follows quite closely the diurnal rhythm in body temperature. This means that efficiency at these tasks is markedly poorer during night watches, when temperature is low, than during day watches, when it is relatively higher. In the traditional "1 in 3" rotating watchkeeping system with four-hour duty spells the worst period from this point of view is the "morning" watch from 0400 to 0800; this watch commences just at the time when body temperature reaches the lowest point in its 24-hour cycle, and is carried out after having been awakened from the middle of a normal night's sleep.

2. During recent observations of UC ratings following this traditional watchkeeping system for long periods on nuclear submarine patrols (Ref 2) it was noted that the members of the watch whose next duty was scheduled for 0400 tended not to go to bed at all beforehand; one object of this practice may have been to avoid the unpleasantness of "turning out" at 0350. However this may be, such a procedure carries the possible consequence of adverse effects on efficiency during the duty spell resulting from loss of an entire night's sleep (Ref 3). This might be particularly important in situations where additional duties prevent ratings from going to bed during the previous afternoon in order to acquire a "sleep-bank".

3. The present experiment was therefore mounted to determine whether the level of efficiency at a "mental" task carried out during the morning watch was significantly affected by whether or not at least some sleep had been taken in the immediately preceding hours of the night. The task was one of "vigilance", involving the detection of target "signals" emanating from a simulated auditory sonar device. It demanded concentrated attention, and was performed continuously throughout the four-hour period, without benefit of any rest-breaks or changes of function. This requirement for unbroken concentration was deliberate, and served two purposes. First, it maximised the stress on the operator and thus increased the likelihood that any effects of the prior loss of sleep would be revealed; and second, it provided a means of obtaining data on performance changes during watch-keeping that, unlike those observed in previous studies, would be free of the possibly confounding effects of task-interruption or job-rotation.

MAIN EXPERIMENT

Method

4. Design and procedure. 12 ratings assigned to APU for training and endurance tests were used as subjects; they were dealt with in two groups of six, each of which was treated exactly alike. The experiment was spread over $2\frac{1}{2}$ of the 6 weeks that each group spent at APU. During the latter part of Week 1 the subjects received extensive training at the sonar task; at the end of this training period they were thoroughly familiar both with the nature of the target signals and the procedure to be adopted in reporting them. During Week 2 the subjects were under "experimental orders" from 2100 Sunday to 1500 Thursday. They were at the APU laboratory from 1945 Monday to 0815 Tuesday, and from 1945 Wednesday to 0815 Thursday. During the daylight hours of Monday and Wednesday they carried out light duties only under supervision at their quarters in the nearby Royal Naval Unit, having had a full night's sleep beforehand. On Tuesday they were given a "make-and-mend". On Thursday they slept until 1500, and then went on leave until 2100 the following Sunday.

5. The design for Week 3 was an exact replication of that for Week 2. The detailed procedure followed while the subjects were at the laboratory each week is described below.

6. On arrival at the laboratory at 1945 on the Monday the subjects rested until 1955. Oral temperatures were then recorded by clinical thermometers, and at 2000 the first of two sonar watches commenced (this first watch served as a "control" period of performance measurement, the purposes of which are described below). At 2100 the subjects were served with de-caffeinated coffee, but were not permitted to break off the task, which continued until 2400. At this time oral temperatures were again recorded, following which three of the subjects turned in on bunks in a nearby room. The remaining three subjects played card games, read, or listened to the radio under continuous supervision until 0350. At this time the three sleeping subjects were roused, and at 0355 oral temperatures were recorded from all subjects. At 0400 the subjects commenced the second watch, which continued until 0800 (decaffeinated coffee being served "on the job" at 0500). At 0800 oral temperatures were again recorded, and the subjects returned to their quarters.

7. On the Wednesday, the procedure followed was exactly the same as for Monday, except that at 2400 the three subjects who turned in were the three who had stayed awake on the Monday night.

8. Although the main interest in these trials was the comparison of the performance levels achieved in the morning watch with and without preceding sleep (hereafter referred to as the "sleep" and "awake" conditions respectively), the inclusion of a prior "control" watch from 2000 to 2400 was considered desirable for three reasons. First, by providing a period of close supervision during these "evening" hours, it solved the practical problem of ensuring that the actual number of hours available for sleep immediately before the morning watch was the same for each subject. Second, it yielded base-line data against which the <u>overall</u> level of performance in the morning watch could be evaluated, ie, it enabled an assessment of the magnitude of the underlying "time of day" effects to be made, for comparison with those observed in previous trials. Third, it produced a situation which paralleled those phases of the traditional watchkeeping cycle in which a period of only four hours off-duty occurs between successive watches, ie, it tested the "worst case" in terms of possible cumulative fatigue effects. This control watch will be referred to in what follows as the "evening" watch.

9. <u>The Task.</u> Subjects performed the signal detection task in individual sound-proofed booths measuring 3 ft x 3 ft x 7 ft; they were thus completely isolated from each other. The simulated sonar output consisted of a three-second burst of amplified thermal ("white") noise at a sound-level of approximately 70 dB, presented over headsets. The outputs alternated with three-second periods of silence, the sequence being intended to represent the operation of successive "sweeps" of the sonar equipment giving rise to a regularly repeated "return". The presence of a "target" signal was indicated by the addition of a 950 Hz tone to the noise for a period of 0.5 seconds commencing one second before the end of the burst. The occurrence of such a tone in the sequence of noise bursts was randomised over the watch with a probability of 0.1 that one would be contained in any particular return. The subjects knew this probability, but were also made fully aware of the fact that the actual number of signals in any particular time period could vary considerably. By this means it was hoped to dissuade them from self-monitoring their performance by counting the reports they made.

10. On detecting the signal, the subject was required to report it as quickly as possible by pressing one of three response buttons placed in a row on a small ledge immediately in front of the chair in which he sat. These three buttons were labelled "maybe", "fairly sure", and "certain" respectively. The subject had to press that button which corresponded most closely to the degree of confidence he felt in his perception of the signal. All responses (correct or incorrect) were recorded; the latency of each correct detection was automatically measured by an electronic timer to an accuracy of 0.01

seconds (any response made within 2.9 seconds of a signal presentation was scored as a correct detection).

11. During the training period, the amplitude of the signal tone was gradually reduced until the mean proportion of signals detected for each group of 6 subjects lay between 50% and 60%., and a stable level of performance had been recorded on several successive sessions. Although the subjects had full "feedback" during this training period, once the trials proper commenced they were given no information about the efficiency of their performance until the entire series was completed.

Results

12. Body Temperature. The means of the body temperatures recorded at the four sampling times are given in Table 1. These means are based on data from both weeks of the experiment; for the readings at 0400 and 0800 they have been calculated separately for the prior conditions of sleep or wakefulness.

13. The changes in temperature observed between 2000 and 0800 correspond closely to those recorded over the same part of the 24-hour period in earlier studies of the normal diurnal variation in this function (Ref 1). Thus temperature declined throughout the evening watch, decreased further during the off-duty period to its predicted low point at 0400, and then rose during the morning watch. The temperature immediately after sleeping was, as might be expected, lower than at the same time after staying awake; however, the difference, although statistically significant ($p < 0.05$) was only some 0.2^{o}F.

14. "Basic" performance measures. The following "basic" scores from the sonar task were assessed for each 30-minute period on watch:

1) Detection rate
2) False report rate
3) Mean latency of responses to signals

Since inspection of the data for the first two of the above scores revealed that there was essentially no difference in the results obtained in the first and second weeks of the experiment, the scores were collated over both weeks. A comparison was then made of the mean trends in these scores during the morning watch under "sleep" and "awake" conditions; these trends were, in turn, compared with those observed during the evening watch.

15. Detection rate. The mean percentage of signals detected in each successive 30-minute period of the morning watch is shown in Figure 1 for the two conditions, together with the corresponding scores during the evening watch.

16. It will be seen that, during the evening watch, the mean detection rate fell from 54.7% in the first 30 minutes to only 35.2% in the last 30 minutes; this substantial decrement was highly significant statistically ($p < 0.01$). The rate of decline was most rapid over the first 90 minutes; ie, the function describing the change in detection rate over time can be defined roughly as "negatively decelerated". Such a function is typically found with "vigilance" tasks of this nature.

17. Whether or not sleep was taken between 2400 and 0400 made no statistically significant difference to the detection rate observed in the first 30 minutes of the morning watch. Note that the mean scores for "sleep" and "awake" conditions in this period (35.8% and 38.1% respectively) were also both very close to that observed in the final 30 minutes of the evening watch, ie, there was no evidence that the four hours rest allowed between the two watches resulted in any recovery of detection efficiency.

18. Over the first two hours of the morning watch the detection rate declined, in both conditions, to an extent comparable with that observed over the corresponding period of the evening watch. There was no significant difference between the conditions in the magnitude of this decline, which was itself statistically significant in both cases ($p < 0.02$); however, from the fourth 30-minute period onwards the detection rate following sleep was slightly but consistently higher than the rate following wakefulness, and performance levels in both conditions remained relatively stable (except for a sharp increase in the final 30 minutes under the "sleep" condition). The mean detection rate over the second two hours of the morning watch was 29.1% in the "sleep" condition, and 22.5% in the "awake" condition; this difference of 6.6% was found to be statistically significant ($p < 0.05$).

19. Thus it would appear that going without sleep beforehand had a definite, though relatively small, adverse effect on detection rate in the second half of the morning watch. Apart from this, the most noticeable aspect of the detection rate score was its marked decline over the night as a whole. This resulted in a substantial lowering of the overall mean score in the morning watch as compared with the evening watch which was statistically significant in both "sleep" and "awake" conditions ($p < 0.01$ in each case).

20. <u>False report rate</u>. The mean percentage of noise bursts responded to incorrectly as containing a signal is shown in Figure 2 as a function of time on watch. It will be seen that, like the detection rate, the false report rate declined during the evening watch, the shape of the curve being very similar to that observed for the former score (ie negatively decerlerated). The effect of this decline was to more than halve the incidence of incorrect responses, the mean percentage dropping from 7.3. in the first 30 minutes to 3.6 in the last 30 minutes, a difference in score which was highly significant statistically ($p < 0.01$). The trend over time was, again, similar to that observed for this measure of performance in other "vigilance" tasks.

21. In the first 30-minute period of the morning watch, the mean false report rate was slightly higher in the "awake" condition than in the "sleep" condition; the difference was statistically significant ($p < 0.05$). Although the actual rates for "sleep" (4.8%) and "awake" (6.3%) were both slightly higher than that observed in the final period of the evening watch, the post-rest increase was found to be statistically significant only in the "awake" condition ($p < 0.01$).

22. The trend in false reports over time during the morning watch was very similar to that shown by correct detections. The rate of incidence declined over approximately the first two hours, after which it remained relatively stable for the remainder of the session. The initial decline was significantly smaller in the "sleep" condition ($p < 0.01$), and from the fourth 30-minute period onwards the false report rate was slightly but consistently higher in this condition than in the "awake" condition. This finding parallels that for detection rate, although the mean difference over the second two hours of the watch (1.1%) was in the present case only marginally significant ($p < 0.10$).

23. Thus it would appear that although going without sleep beforehand resulted in a slight increase in false reports at the start of the morning watch, the subsequent decline in the rate at which these reports were made was clearly more pronounced in this "awake" condition than after sleep, and there was a fair indication that the resultant lowered level of reporting was sustained for the second half of the four-hour period. Whether these findings should be interpreted as beneficial or not clearly depends partly on the operational cost of a false report in the practical situation, and, in any case, should be judged in relation to the accompanying changes in the efficiency with which actual signals were detected.

24. Apart from these relatively small differential effects of the prior sleep conditions

on performance in the morning watch, the main feature of the false report score was, as in the case of correct detections, its marked decline over the night. Once again, this resulted in a lowering of the overall mean score in the morning watch as compared with the evening watch which was statistically significant in both "sleep" and "awake" conditions ($p < 0.05$ and $p < 0.01$ respectively).

25. Mean latency of responses to signals. In the first week of the experiment defects in the timing mechanism reduced the number of response latencies recorded to an extent which prevented assessment of reliable means. Thus the evaluation of changes in this score is based on the readings for the second week only, during which period complete recordings were obtained from 11 of the 12 subjects. Figure 3 shows the overall mean latencies for these 11 subjects as a function of time on watch.

26. It will be seen from Figure 3 that in most respects the trends for response latency were very similar to those for detection rate, viz, an increase in latency (ie , a decline in efficiency) over time during both watches, with the overall average latency in the morning watch being significantly longer than that in the evening watch in both "sleep" and "awake" conditions ($p < 0.01$ and $p < 0.05$ respectively). During the morning watch, however, there was no evidence of the consistent difference between conditions over the second half of the duty period that was seen with the detection rate score. The only apparent consequence of going without sleep beforehand was in the first 30-minute period. At this time the mean latency in the "awake" condition was some 7 percent shorter than in the "sleep" condition, and also some 8 per cent shorter than in the final 30 minutes of the evening watch. Both these differences were statistically significant ($p < 0.05$), and the second of them suggests that some recovery in efficiency took place over the four-hour rest interval when this was spent in wakefulness.

27. In respect of the speed with which detected signals were reported, therefore, performance was, for a short time only, a little worse after sleeping than after remaining awake. It should be noted that this slight relative initial slowness in reaction after being awakened from sleep coincided with a similarly slight relative reduction in the total number of reports made (see previous sections); it is possible that both these findings may be reflections of the same phenomenon, namely, the residual effects of "sudden awakening" such as have been reported elsewhere (ref 4).

28. "Derived" performance measures.

Discrimination efficiency. To determine whether the observed changes in the detection and false report scores reflected alterations in the intrinsic ability to discriminate signals from non-signals the two scores were combined to yield the measure of signal "detectability" ($\underline{d}'$) provided by the theory of signal detection (Ref 5), on the assumption that the model proposed by this theory was in fact valid for the present data (see Ref 6 for a discussion of this issue).

29. Mean values of $\underline{d}'$ are shown in Figure 4 for each 30-minute period of the two watches (note: these values are based on only 11 subjects, since in one case estimates of $\underline{d}'$ could not be made in certain 30-minute periods owing to a total absence of false reports). During the evening watch the mean value of $\underline{d}'$ remained at a relatively constant level (the apparent slight overall decline over the four-hour session was found not to be statistically significant). Thus discrimination efficiency could be said to have been effectively stable over this watch. According to signal detection theory, the implication is that the marked fall-off in the rate at which signals were reported during the session was due primarily to a progressive shift in the "response-decision" criterion to a more "cautious" level. Such a shift has been held to occur with time in many other vigilance tasks.

30. By contrast, during the morning watch there appeared to be, in both "sleep" and "awake" conditions, considerable variation in the mean value of d'. However, this variation was found to be statistically significant only in the "sleep" condition ($p < 0.01$); here there was a sharp decline from the early part of the watch to the middle section, followed by a marked increase which was particularly evident in the final 30-minute period. Explanations for the detailed structure of this rather unusual trend are not immediately apparent, but it is hypothesised that the rise in discrimination efficiency over the second half of the watch may reflect the increase in arousal associated with the rapid rise in body temperature that occurs over this part of its 24-hour cycle (Ref 1). This hypothesis is given some support by the finding that in the final 30-minute period mean d' was significantly lower in the "awake" condition than in the "sleep" condition ($p < 0.01$), a result that might be expected on the not unreasonable assumption that in the former case the improvement due to increasing arousal was being cancelled out by the effects of the prior sleep deprivation.

31. It is of interest to note that the overall mean level of d' in the evening watch was considerably higher than the overall levels observed in the morning watch. The difference between the two watches was statistically significant in both "sleep" and "awake" conditions ($p < 0.05$) and $p < 0.1$ respectively). This implies that the substantial reduction in the rate at which signals were detected in the morning watch as a whole resulted, at least in part, from a deterioration in the intrinsic ability to distinguish signals from non-signals.[1]

32. Confidence level. The relative use of the three response buttons was examined for evidence of any systematic changes in the degree of confidence expressed in making reports. A measure of confidence level was obtained for each subject by assigning a weighting of 1, 2 or 3 to each report according to whether it was made with the response button labelled "maybe", "fairly sure" or "certain" respectively, and then computing the average score per report for each 30-minute period. Figure 5 shows the mean trends in this index during the two watches.

33. Although Figure 5 suggests that some changes occurred in confidence level in successive 30-minute periods of the sessions, this apparent within-watch variation was found not to be statistically significant either in the evening watch or in the morning watch (both conditions). On the other hand, the overall score for the morning watch in the "sleep" condition was significantly lower than the corresponding value for the evening watch ($p < 0.02$); whereas no such significant lowering was found in the "awake" condition. Thus it would seem that staying awake beforehand tended to overcome the slight general reduction in confidence level that took place in the morning watch when this was preceded by sleep.

Discussion

34. Effects of staying awake. The results suggest that the effects of remaining awake before the morning watch instead of turning in for four hours were, although statistically significant, relatively small. It would seem that in the initial 30 minutes of the watch

---

[1] It is possible that this reduction was also being determined to some extent by a shift in the decision criterion to a more "cautious" level; however, since changes in this criterion can only be inferred with confidence where d' remains constant, a definite conclusion on this point cannot be drawn (for the same reason no statement can be advanced concerning the possible contribution that shifts in the decision criterion may have made to the changes in the rate of reporting that occurred as a function of time within the morning watch).

there was some slight advantage to be gained by prior wakefulness, since under this condition faster responses were made to signals when they were detected. However, this advantage tended to be cancelled out by a propensity for more false reports to be made.

35. During the remainder of the first half of the watch there appeared, again, to be some small advantage in having remained awake, since over this time the rate at which false reports were made in this condition declined more sharply than it did after prior sleep; whereas there were no significant differences in this respect for detection rate, or for the speed at which detected signals were reported.

36. By contrast, over the whole of the second half of the watch the advantage in terms of the main score (ie, detection rate) lay with the "prior sleep" condition; this advantage became quite clear in the final 30-minute period and was only partly nullified by a tendency for a few more false reports to be made, and for subjects to respond in a slightly less confident manner.

37. Effects of time of watch and task duration. The most notable feature of the results was the substantial change in the number of both correct and incorrect reports made as a function of time of day and of the duration of the watch. During the evening watch both of these latter factors would seem to have combined to produce a marked decline in detection rate over the period from 2000 to 2400. In the morning watch, on the other hand, the two factors appear to have partially cancelled each other out, in that, although there was a further reduction in detection rate in the watch taken as a whole (which may have resulted, at least partly, from a lowering of the intrinsic ability to distinguish signals from non-signals), the within-watch decrement was less pronounced, presumably because over the second half of this watch the "time of day" effect was working in the opposite direction to that of task duration.

38. The relative magnitudes of the changes in the various performance measures can be readily compared by inspection of Figure 6, where in each case the scores for successive 30-minute periods of the two watches have been plotted as a percentage of the initial "base-line" score recorded in the first 30-minute period of the evening session (note: in the morning watch the values shown are for the "sleep" condition).

39. It will be seen that the progressive decline over time in the primary score of detection rate that led to a low-point in performance between 0600 and 0730 was so marked that over the latter period the level of efficiency was less than 50 per cent of the base-line value. This pronounced decrement was accompanied by an even greater proportional fall in the false report rate. The third "basic" response measure, reporting speed, showed a similarly progressive, but rather smaller proportional degradation that resulted in a relative reduction in performance level in the morning watch comparable in magnitude to that exhibited by the derived index of discrimination efficiency, $\underline{d}'$. Least relative decline was noticeable in the measure of confidence level.

40. The fact that the deterioration in detection rate was so much more pronounced than was previously observed over the same times of day in the earlier experiments on watchkeeping in which job-rotation was routinely practised (Ref 1) suggests that the latter procedure tends to mitigate the influence of diurnal rhythms on efficiency, and that the present results, based as they are on a task situation deliberately designed to impose maximal stress in terms of both continuous performance within a watch and minimal rest interval between watches, may well provide a truer estimate of the magnitude of the "time of day" effects that could be expected in fully alerted operational conditions over the night hours.

41. Watchkeeping is, by definition, a "round-the-clock" operation. If the conclusion in the preceding paragraph is correct, it becomes necessary to determine the effect that the present "stress" task-situation has on performance efficiency during that part of the 24-hour period not covered by the experiment just described, ie, during "daytime" hours, when sleep deprivation is not a confounding issue. In order to do this a further trial was carried out to assess the changes in performance that occur when subjects carry out the same watch routine during "normal" hours, after an uninterrupted night's sleep beforehand. The results of this "control" experiment are reported below.

CONTROL EXPERIMENT

Method

42. Design and Procedure. Eleven fresh ratings were used as subjects for this experiment; they were dealt with in two groups, one of five, and one of six in number. The procedure was essentially the same as that followed for the main experiment, with the exceptions that the duration of the trial and the actual times of day and week when testing was carried out were altered as follows:

1. The times of the watches were from 0800 to 1200 (forenoon watch) and from 1600 to 2000 (dog watch).

2. The whole experiment was conducted over a period of 2 rather than $2\frac{1}{2}$ weeks.

3. The training was accomplished on the Monday and Tuesday of the first week, rather than during the latter half of the preceding week.

4. In the first week, the tests were conducted on Wednesday and Friday; and in the second week on Tuesday and Thursday; as compared with the "Monday night and Wednesday night" routine followed in both weeks of the main experiment.

During the first week, the subjects were under "experimental orders" from 2100 Tuesday to 0700 Saturday. A "make-and-mend" was given on the Thursday. After leave commencing at 0800 Saturday, the subjects were put under "experimental orders" again from 2100 Monday to 0700 Friday, a "make-and-mend" being given on the Wednesday.

43. At the laboratory, the detailed procedure followed during each watch was exactly the same as that followed in the main experiment. However, in the four-hour period between watches the subjects were allowed to return to their quarters for lunch; after the meal they were kept under supervision in these quarters until returning to the laboratory for the second watch.

44. The task. This was identical to that used in the main experiment.

Results

45. Body Temperature. The means of the body temperatures recorded at the four sampling times are given in Table 2 (as before, these means are based on readings taken from both weeks of the experiment). Apart from the fact that the mean reading at 0800 was slightly higher than expected for this time, during the day body temperature rose by an amount which corresponds quite closely to that recorded over the same part of the 24-hour period in the earlier studies already cited (Ref 1).

46. "Basic" Performance measures. Before attempting any comparison of performance levels in the "forenoon" and "dog" watches, average scores for detection rate, false report rate, and mean latency of responses to signals were computed separately for each of the eight watches held during the experiment, in order to determine whether any systematic long-term "practice" trends were observable in these measures over the two-week period as a whole. If such trends were found it would be necessary to make some adjustment to the results before comparing forenoon and dog watches, since performance levels at the two times of day would inevitably be confounded by the practice effect (note: a similar adjustment would also then be required to be made to the results of the main experiment, the practice effect in which was not readily assessable directly, but which would be assumed to have been the same as in the present case).

47. The average scores for the eight watches are shown in Table 3, from which the following conclusions can be drawn:

1. There is no evidence of any systematic practice effect over the two-week period in the scores of either detection rate or response latency.

2. Although the false report scores were lower in the second week than in the first, there is no evidence that this reflected a systematic trend operating throughout the experiment. A slight increase in this score occurred during the second week, but the changes were so small that they can safely be ignored for present purposes.

Since no adjustment for practice effect appeared to be required, scores were collated over the four trial days. An examination was then made of the mean trends in these scores during the forenoon and dog watches.

48. Detection rate. *The mean percentage of signals detected in each successive 30-minute period of the two watches is shown in Figure 7.* The decrement in performance during each watch was essentially similar in nature to that observed during the evening watch in the main experiment, although its eventual extent was slightly less (approximately 14 percentage points), and more in line with that usually found in vigilance tasks conducted during "normal" hours. The decline in score from the first to the last 30 minutes was of almost the same magnitude in both forenoon and dog watches, and highly significant statistically in each case ($p < 0.01$).

49. As the graphs suggest, there was very little difference between the overall mean scores in the two watches (forenoon 49.2%; dog, 47.3%) and this difference was not significant statistically.

50. False report rate. The mean percentage of noise bursts responded to incorrectly as containing a signal is shown as a function of time on watch in Figure 8, which shows that the incidence of incorrect responses almost halved during both forenoon and dog watches, as it did in the evening watch. The "low-point" was reached in the penultimate 30-minute period. The fall from the first 30-minute period to the latter was of the same magnitude in both watches, and was highly significant statistically in each case ($p < 0.01$). There is a suggestion in the graphs that the decline in the false report score occurred earlier during the forenoon watch than during the dog watch, but on analysis the apparent difference in the extent of the fall in the rate from the first to the second 30-minute period was found not to be statistically significant.

51. As is suggested by inspection, there was negligible difference between the overall mean scores in the two watches; in fact, the difference was less than 0.1 percentage points (forenoon, 1.68%; dog, 1.74%) and was not statistically significant.

52. Mean latency of responses to signals. The overall mean latency of responses made to signals as a function of time on watch is shown in Figure 9, from which it will be seen that the trends in response latency during each watch were, once again, similar to that observed during the evening watch. As in the latter case, latency increased (ie efficiency declined) over time, and in much the same way as did detection rate. The extent of the increase as in the case of the detection rate score, was somewhat smaller in the present experiment. The increase was virtually identical in forenoon and dog watches, and the change from the first to the last 30-minute period (representing a slowing in reporting speed of some 10 per cent) was highly significant statistically in each case ($p < 0.01$).

53. It would appear from the graphs that there was virtually no difference between the overall levels of performance in the two watches; in fact, the actual difference was less than 0.01 seconds, and was not statistically significant.

54. "Derived" performance measures. Table 3 shows that, as in the case of the "basic" scores, there were no systematic practice effects in either of the derived measures over the two-week period of the experiment. Scores for these measures were therefore collated over the four trial days, in the same way as before.

55. Discrimination efficiency. Mean values of d' are shown as a function of time on watch in Figure 10, where the graphs appear to suggest some degree of variation in the score in successive 30-minute periods of the sessions. However, this variation was found not to be statistically significant in either the forenoon or the dog watches. Thus discrimination efficiency could be said to have remained effectively stable during both watches, as it did in the evening watch. The implication of this in terms of signal detection theory is the same as that drawn in the latter case, namely that the marked fall-off in the rate at which signals were reported during each watch was due primarily to a progressive shift in the response-decision criterion to a more cautious level.

56. The indication in Figure 10 is that the overall mean value of d' differed very little in the two watches; in fact, the actual difference was less than 0.1 unit (forenoon, 2.27; dog, 2.21), and was not statistically significant.

Note: the overall average values of d' observed in the present experiment were somewhat higher than those recorded in the main experiment. This was a necessary computational result of the fact that, whereas, in general, detection rates were very similar in the two experiments, the overall incidence of false reports was considerably lower in the present case. Thus the only "real" difference between the two experiments in terms of general levels of performance was in this particular score of commissive errors. However, it is not considered that this difference invalidates comparisons between the experiments, since (a) these comparisons are to be made in terms of changes over time rather than in absolute levels and (b) in any case, a difference in mean false report rate of the magnitude concerned is not uncommonly found between successive groups of subjects given the same vigilance task, whatever its nature. These differences occur because (for reasons which are not in the present state of knowledge fully understood) variation in false report rate between individuals is always very substantial, and in samples of the size used in the present investigation (ie N's of 12 and 11), it is to be expected that occasionally one sample will contain more of the "high scorers" than another. This would appear to have happened here.

Confidence level. Figure 11 shows the mean trends in the index of confidence expressed in making reports, during each of the watches. At first sight, the graphs suggest that confidence rose to some extent within each watch, and in a reasonably systematic manner. However, analysis showed that, as in the two watches of the main experiment, the overall variation between successive 30-minute periods was not statistically significant in either the forenoon or the dog watches. It must be concluded, therefore, that there is no sound evidence for any change in this score as a function of time during a four hour watch, whether the latter is held during the day or the night. (Note: examination of the records revealed that individual differences were considerable in both experiments, and that the apparent rise in confidence during the session indicated by the trend in the mean scores in all but the morning watch reflected the performance of only a minority of subjects in each case; investigation of the characteristics of this minority might prove a subject worthy of further research).

57. A further suggestion made by the graphs in Figure 11 is that the overall mean confidence score rose between watches. However, comparison of the actual mean values recorded (forenoon, 1.89; dog, 1.91) showed that the difference of 0.02 units was not statistically significant.

Discussion

58. Within-watch variation in efficiency. Since there was no significant difference between forenoon and dog watches in the manner in which the various scores changed as a function of time within the session, the results for the three basic scores (which were also the only ones in which these changes were statistically significant) were collated over both watches. In order to enable direct comparisons to be made with the results of the main experiment, these collated scores for successive 30-minute periods of the session are plotted in Figure 12 as a percentage of the initial base-line score recorded in the first of these periods (as in Figure 6).

It will be seen that the relative magnitude of the within-session changes in the three scores was the same in the present experiment as in the evening ("control") watch of the main experiment, ie, the greatest change occurred in the false report score, the smallest in the latency of responses to signals. In terms of the actual extent of the proportional variation, the fall-off in false report rate (to approximately 60% of its initial value) as the watch progressed was of essentially the same magnitude as observed in the evening watch. The proportional decrements in detection rate and response latency were both slightly smaller than seen in the evening watch, confirming the earlier conclusion that, in that watch, time of day and time-on-task effects were combining to enhance the decrements somewhat. In general, however, the trends were very similar in both experiments, and demonstrate once again the substantial size of the changes in performance to be expected during a continuous "vigilance" watch even when this is held during "normal" (ie, waking) hours, and where there is no question of any sleep deprivation or other related stress affecting efficiency.

59. Performance over the day. The absence of statistically significant change in the mean levels of any of the scores between forenoon and dog watches was an outstanding feature of the control experiment. The chief importance of this finding is that it demonstrates the absence of any obvious "cumulative fatigue" effects arising from the extended hours of duty entailed in the 4-ON, 4-OFF, 4-ON schedule followed. At the same time, however, the results are different from those which might have been expected on the basis of the results of the earlier study of four-hour duty watches previously mentioned (Ref 1). In that study certain overall performance levels in a similar task did differ between the relevant watches. In terms of proportional change, detection rate was some 15% higher in the dog watch than in the forenoon watch, and response latency was about 7% shorter (no change was observed in the false report score, and the two "derived" measures of performance were not computed). In view of these sizeable improvements in efficiency, it must be surmised that the differences in the routine followed in the present experiment were sufficient to negate them.

from the present one. First, jobs were rotated within each watch, so that only half (two unbroken hours) of each watch was actually spent on the simulated sonar task. Secondly, the watchkeeping schedule adhered to was such that the subjects had always rested for eight hours immediately prior to keeping the dog watch, which was not, as in the present case, held on the same day as the forenoon watch. The absence of job rotation combined with the shorter length of the inter-watch rest interval in the present experiment may well have had the effect of generating an amount of "fatigue" which, while just sufficient to overcome the improvements in efficiency which might otherwise have occurred, was not large enough to produce an actual deterioration in performance over the day.

61. Relevance to the results of the main experiment. If the conclusion in the previous section is correct, it should be possible, by comparing the findings of the main experiment with the relevant results from the earlier study, to estimate the extent to which the observed changes in mean performance levels between the evening and morning watches were enhanced by the "fatigue" assumed to have been induced by the more rigorous schedule followed in the present case. Table 4 summarises the proportional changes in the various scores observed in the two investigations.

62. It will be seen that, in the case of detection rate, the deterioration in score in the present experiment was over $2\frac{1}{2}$ times as great under "sleep" conditions, and over 3 times as great under "awake" conditions, as in the earlier study. Clearly, fatigue contributed substantially to the observed changes in this measure, either by adding to, or interacting with, the predicted change due to diurnal rhythm per se (note: in the earlier study the morning watch, like the dog watch, was preceded by a long rest). This effect was even more dramatic for the false report score, where it would appear that the whole of the change was attributable to the particular routine followed in the present experiment (it is of interest to note that this experiment is the only one of the entire watchkeeping series carried out at APU (Ref 1) in which any statistically significant change in this score between watches held at different times of day has ever been observed).

63. In contrast to the above results, the extent to which the latency of responses increased over the night would appear to have been little affected by fatigue factors. Although this might seem rather surpirsing, it is explicable if it is hypothesised that the marked enhancement of the reduction in signal - reporting rates noted above was due mainly to a rise in the incidence of intermittent failures to perform the task at all, and only partly (if at all) to an increase in the underlying degradation of ability. If it is surmised that these failures were due to periods of excessive drowsiness, or even of actual sleep, then it seems not unreasonable to suppose that at other times, when the subjects were actively engaged on the task, signals would have been reported at a speed which was much the same as "normal" for that time of day and for the particular time for which the watch had continued.

64. This explanation must, of course, remain speculative. The fact that there was a 10 percent reduction in the index of discrimination efficiency, $\underline{d}'$, suggests that the underlying ability to distinguish signals from non-signals was, in fact, significantly lowered in the morning watch. However, this index was calculated over relatively gross time periods; it might possibly have been the case that discrimination efficiency when awake during these periods was actually unaffected. A test of this hypothesis could only be made by further experiments in which wakefulness was monitored independently (eg, by EEG) throughout the watches.

CONCLUSIONS

65. Statistically significant changes in performance at a signal detection task carried out continuously during the morning (0400 - 0800) watch are produced by staying awake rather than turning in for four hours immediately beforehand. Different aspects of performance are affected in different ways and at different times within the watch, but the most important finding is probably that detection rates are consistently lower during the second half of the watch when operators have had no prior sleep.

66. Even when four hours sleep is taken beforehand, levels of performance in the morning watch are strikingly different from those obtained during a watch held from 2000 to 2400 the previous evening. The most obviously important finding here is that for the watch as a whole detection rate is reduced by some 30 percent, and during the period from 0600 to 0730 the rate is less than 50 percent of that recorded at the start of the evening watch when personnel are fresh. These very substantial degradations in correct detection rate are accompanied by even greater falls in the incidence of false reports of signals. Although by itself this latter finding might be considered advantageous to the optimal functioning of a detection system, in the present context it is interpreted as meaning that the operators are failing to maintain adequate attention to the task.

67. To some extent this failure would seem to be due to fatigue arising from the continuous nature of the task and the rigorous 4-ON, 4-OFF, 4-ON watchkeeping schedule followed in the present experiment. This fatigue appears to enhance the underlying effect of diurnal rhythm which, by itself, is known to produce a substantial lowering of efficiency in the morning watch.

68. The results also show, once again, that efficiency at prolonged uninterrupted vigilance tasks declines considerably simply as a function of time during any watch, even when sleep-deprivation is not a confounding factor.

69. Mitigation of both within-watch and between-watch decrements in efficiency at detection tasks could be achieved by interrupting the task at regular intervals, or by practising some form of systematic job rotation. Further alleviation of between-watch decrement over the night hours could be obtained by allowing a period of at least eight hours rest between successive watches. However, such measures will not prevent degradation of performance in the morning watch from occurring altogether, since this is a natural concomitant of the diurnal rhythm. Completely stable levels of efficiency over the night watches can only be expected if certain personnel are permanently assigned to standing night duty throughout the duration of a patrol, in order to allow physiological adaptation to such unusual hours of work to occur. Re-arrangement of watchkeeping schedules to encompass this end is strongly recommended.

ACKNOWLEDGEMENTS

Many people helped make this work possible, not least the experimental subjects and the staff of the Royal Naval Unit, Cambridge. On the civilian side, we would like to express our gratitude to the following individuals for their valuable assistance with the administrative, tehnical or statistical aspects of the study: A K Copeman, A Craig, J L Dennett, S Folkard, M V Lee, B J Morgan, M I Nimmon-Smith, R G Olsson, K Tayler and L F Thomson.

REFERENCES

1. OES 513 Experimental studies of watchkeeping - summary report, by W P Colquhoun, M J F Blake and R S Edwards. October 1969. (Published as "Experimental studies of shift work", in Studia Laboris et Salutis, 1969, 4, pp 47 - 64).

2. SMS 1/72 Watchkeeping studies in HMS REPULSE - interim report, by W P Colquhoun and M W P H Paine. January, 1972.

3. RNP 58/913 The effects of loss of sleep on performance, by R T Wilkinson, 1958. (see also "Sleep deprivation", by the same author, in Edholm O G, and Bacharach A L (Eds): The Physiology of Human Survival, New York: Academic Press, 1965).

4. OES 523 Performance after awakening at different times of night, by R T Wilkinson and M Stretton. December 1970. (Published in Psychonomic Science, 1971, 23, pp 283-285).

5. Green, D M and Swets, J A Signal Detection Theory and Psychophysics. New York: Wiley 1966.

6. Broadbent, D E Decision and Stress. London: Academic Press, 1971.

TABLE 1

Mean body temperatures (^{o}F) at the beginning and the end of the two watches in the main experiment

Evening Watch		Morning Watch		
2000	2400		0400	0800
		Sleep:	96.74	97.46
97.87	97.35			
		Awake:	96.96	97.38

TABLE 2

Mean body temperatures (^{o}F) at the beginning and the end of the two watches in the control experiment

0800	1200	1600	2000
97.80	97.85	98.22	98.23

TABLE 3

Mean performance scores in each of the 8 watches of the control experiment

	WEEK 1				WEEK 2			
	Day 1		Day 2		Day 1		Day 2	
	Forenoon	Dog	Forenoon	Dog	Forenoon	Dog	Forenoon	Dog
Detection Rate (%)	49.8	47.8	50.3	50.1	49.1	43.0	47.7	48.6
False Report Rate (%)	2.1	2.1	2.0	2.0	1.1	1.3	1.5	1.6
Response Latency (sec)	1.4	1.4	1.4	1.4	1.5	1.5	1.5	1.5
d'	2.2	2.1	2.2	2.2	2.4	2.3	2.3	2.2
Confidence Score	1.9	1.9	1.9	1.9	1.9	1.9	1.9	2.0

TABLE 4

Proportional changes in overall mean scores between evening and morning watches in the present experiment and in the earlier study

	Earlier Study	Present Experiment "Sleep"	Present Experiment "Awake"
Detection Rate	- 11.6%	- 29.4%	- 37.6%
False Report Rate	(0) *	- 32.5%	- 42.3%
Response latency	+ 6.6%	+ 8.0%	+ 7.2%
Discrimination Efficiency	-	- 10.1%	- 13.2%
Confidence level	-	- 3.0%	(0)*

*Change not significantly different, therefore assumed to be zero.

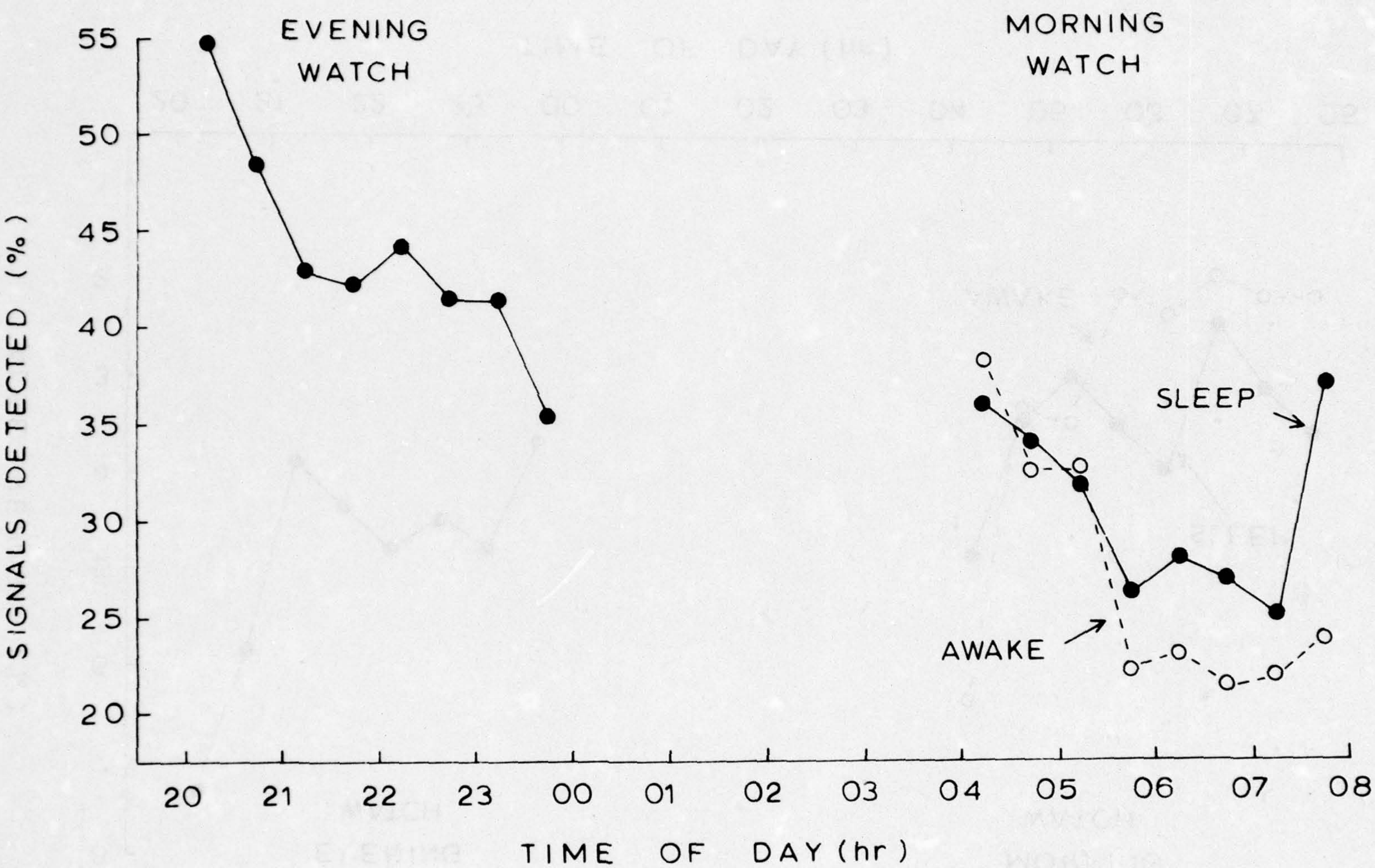


Fig. 1. Main experiment: mean percent signals detected in successive 30-minute periods of the watches.

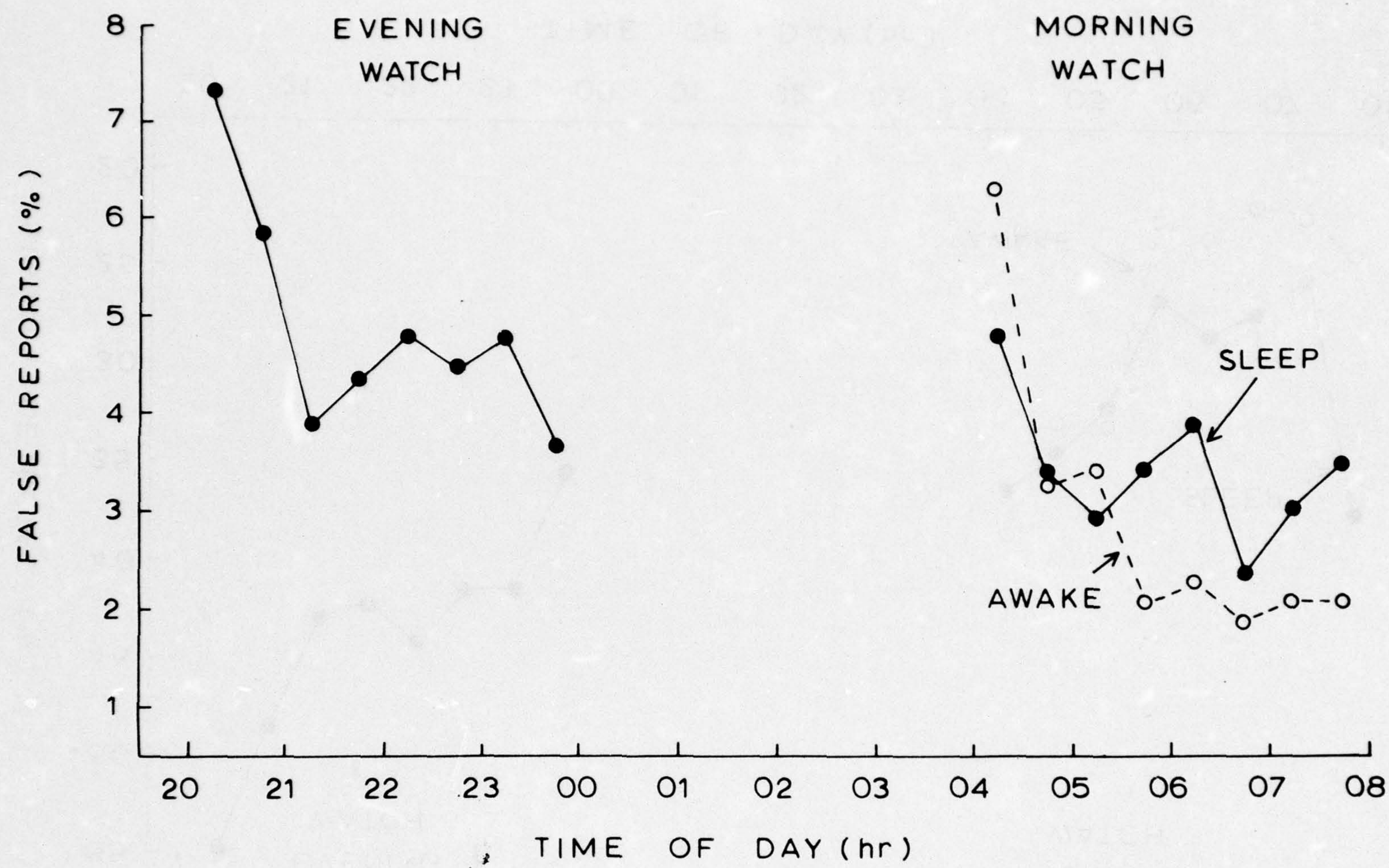


Fig. 2. Main experiment: mean percent false reports made in successive 30-minute periods of the watches.

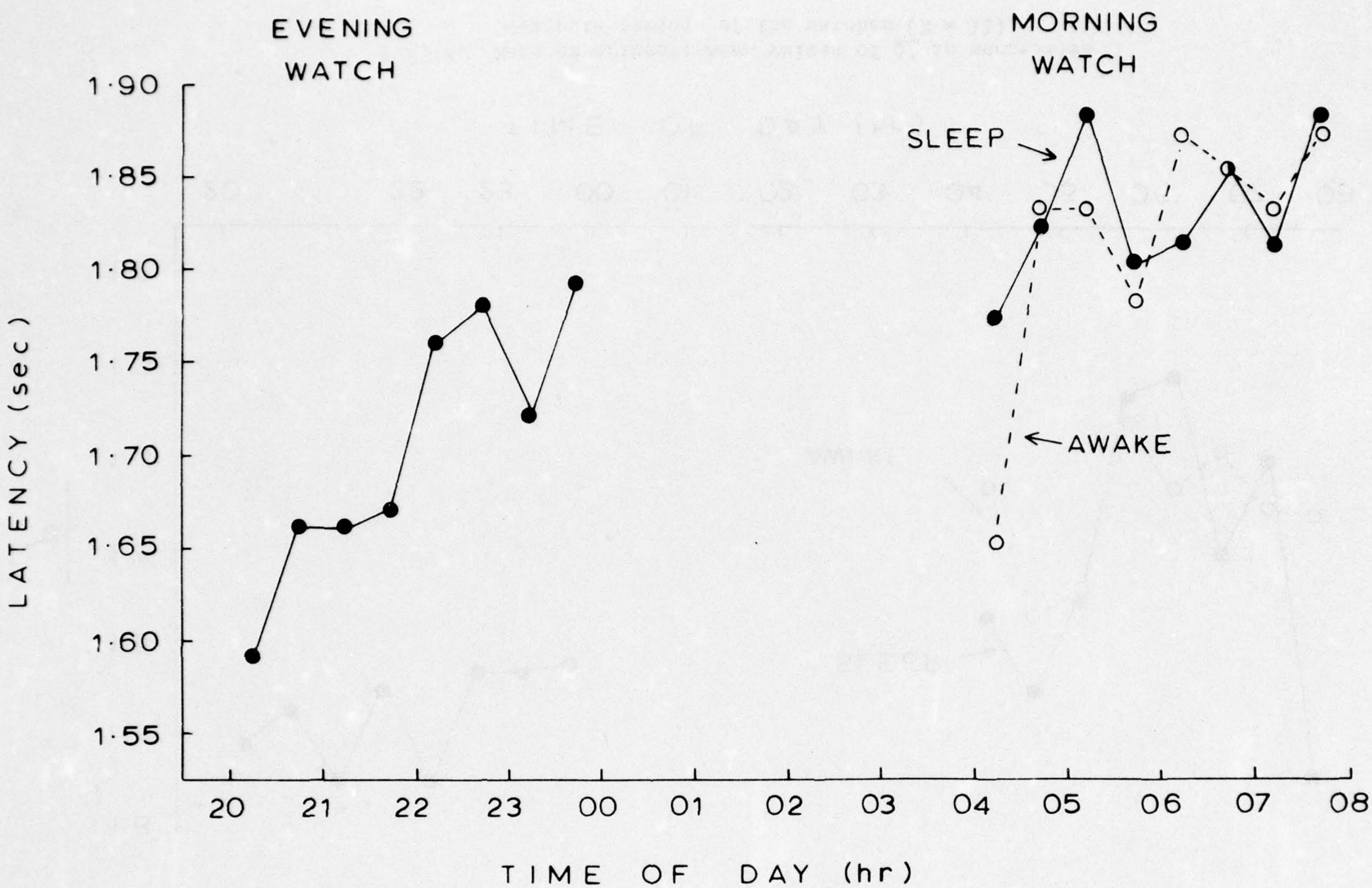


Fig. 3. Main experiment: mean latency of responses to signals in successive 30-minute periods of the watches (N = 11; results from Week 2 only).

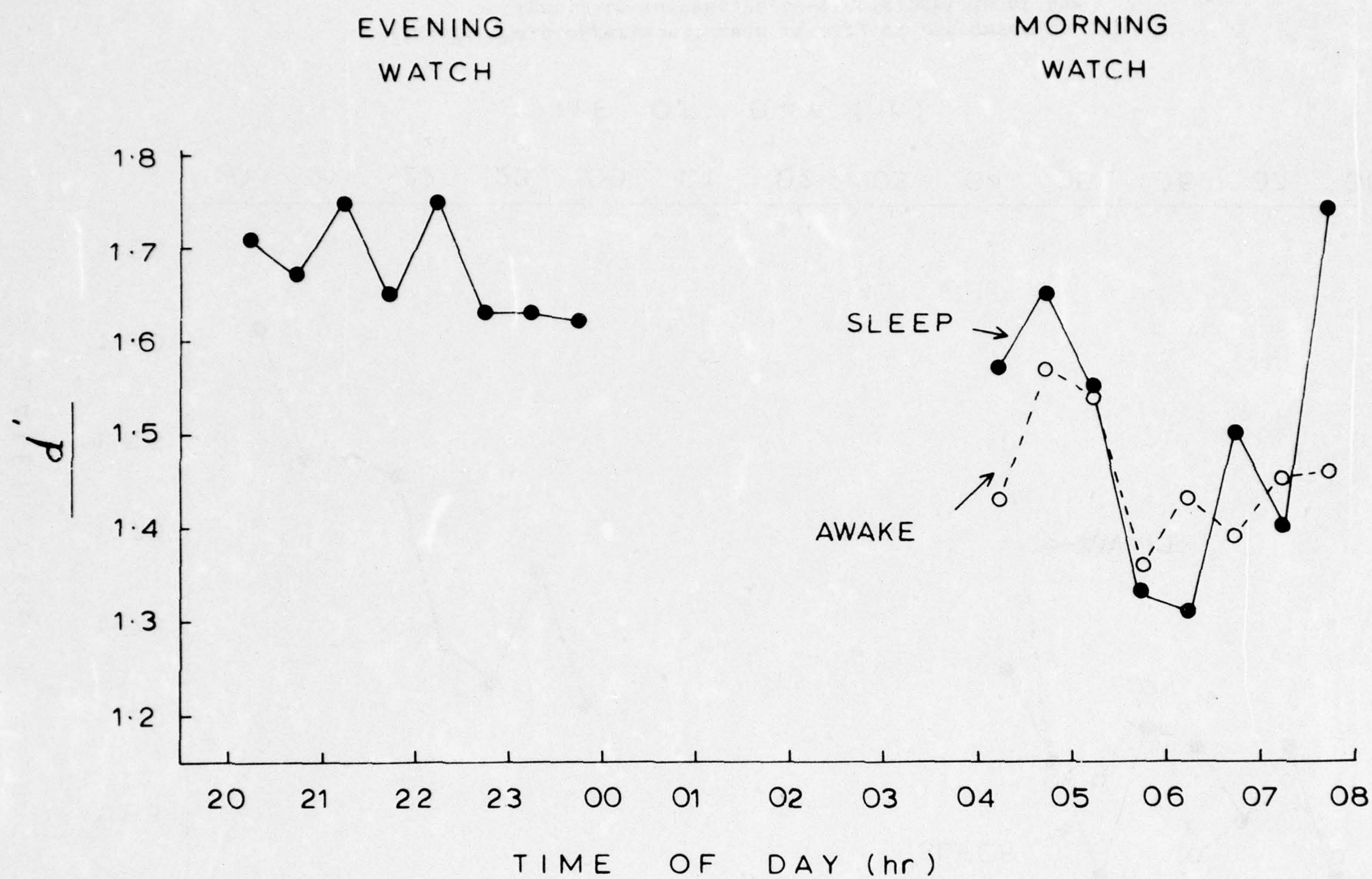


Fig. 4. Main experiment: mean values of d' in successive 30-minute periods of the watches (N = 11).

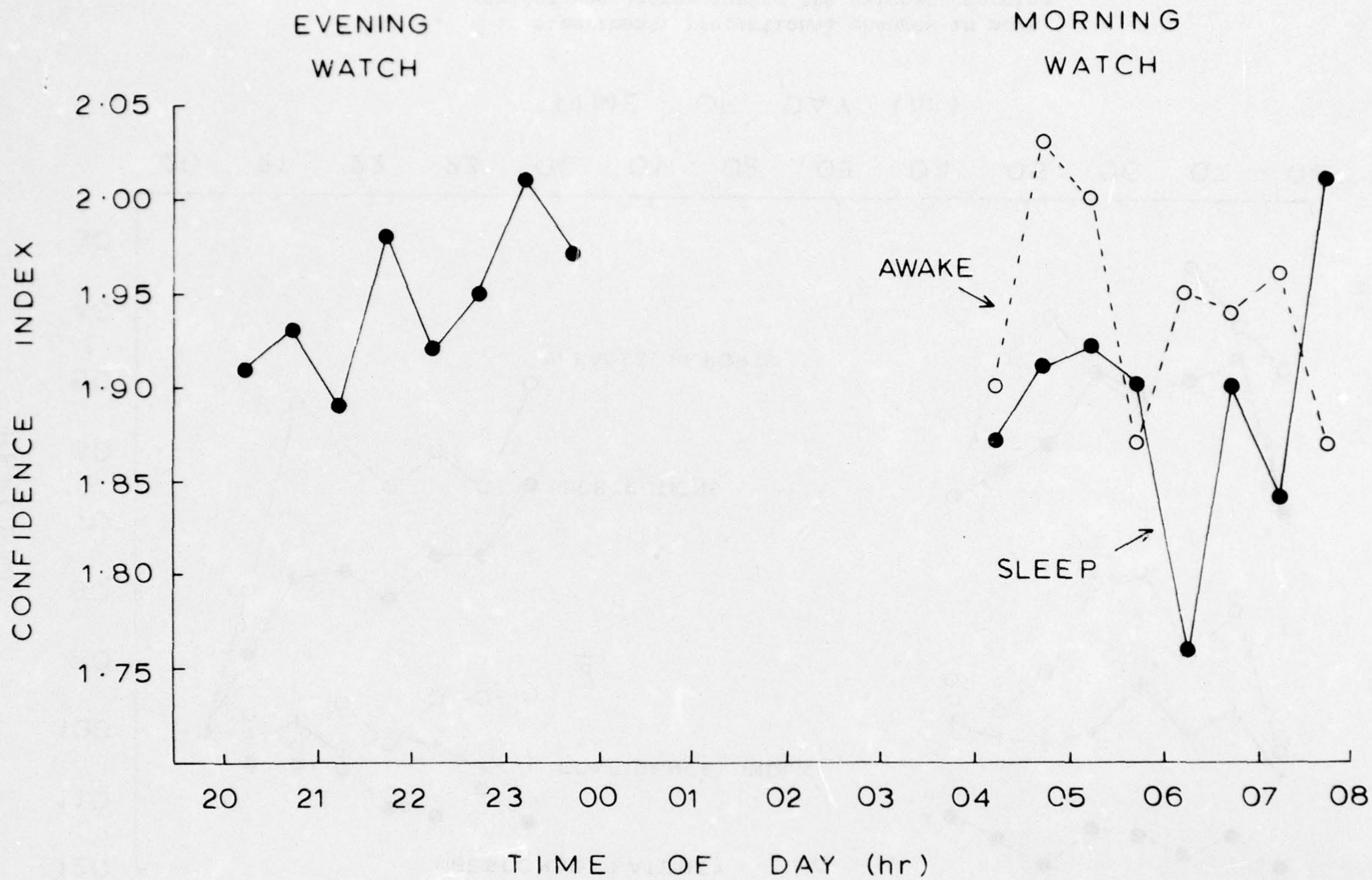


Fig. 5. Main experiment: mean "confidence score" in successive 30-minute periods of the watches.

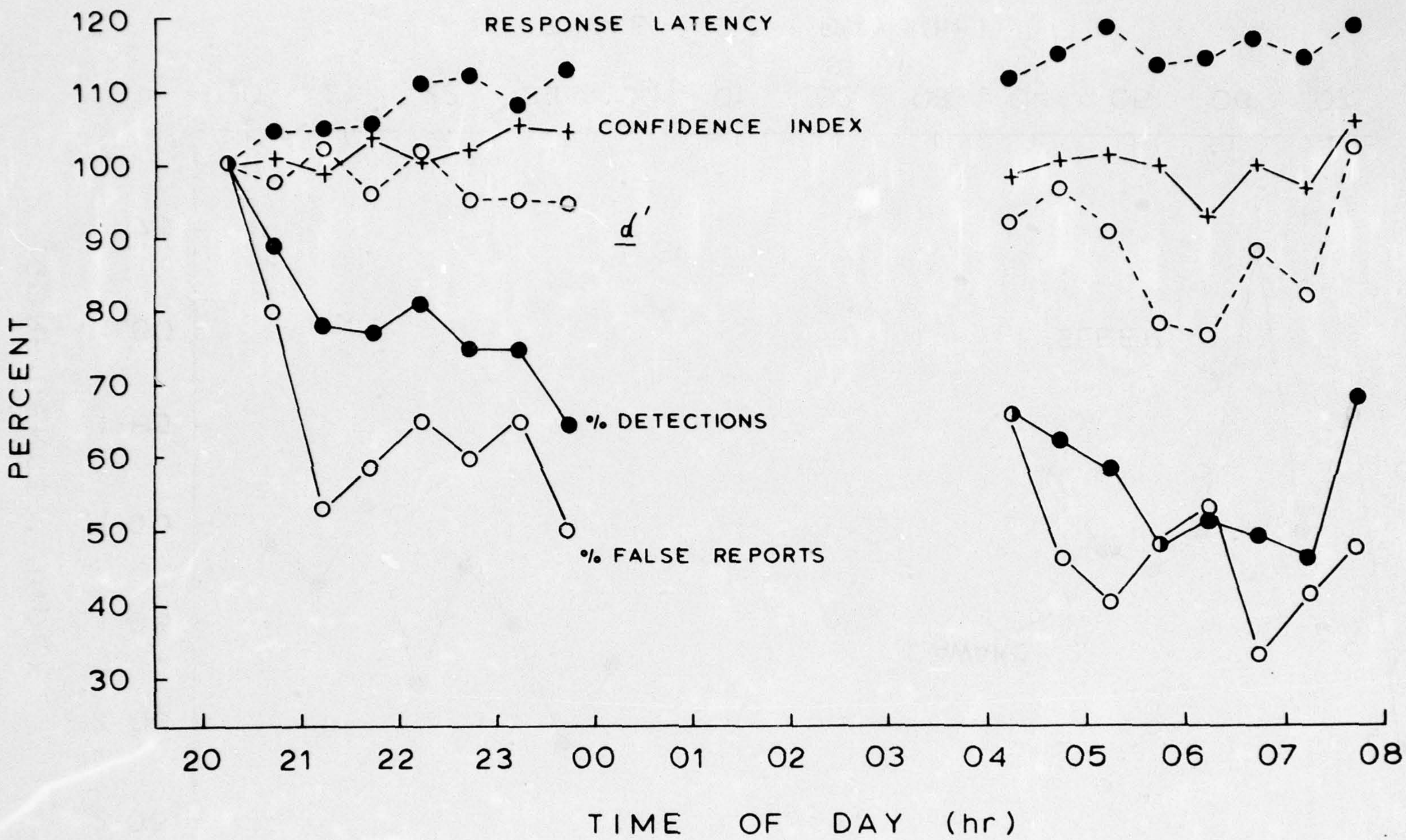


Fig. 6. Main experiment: proportional changes in mean performance scores during the watches (morning watch values shown are for the "sleep" condition).

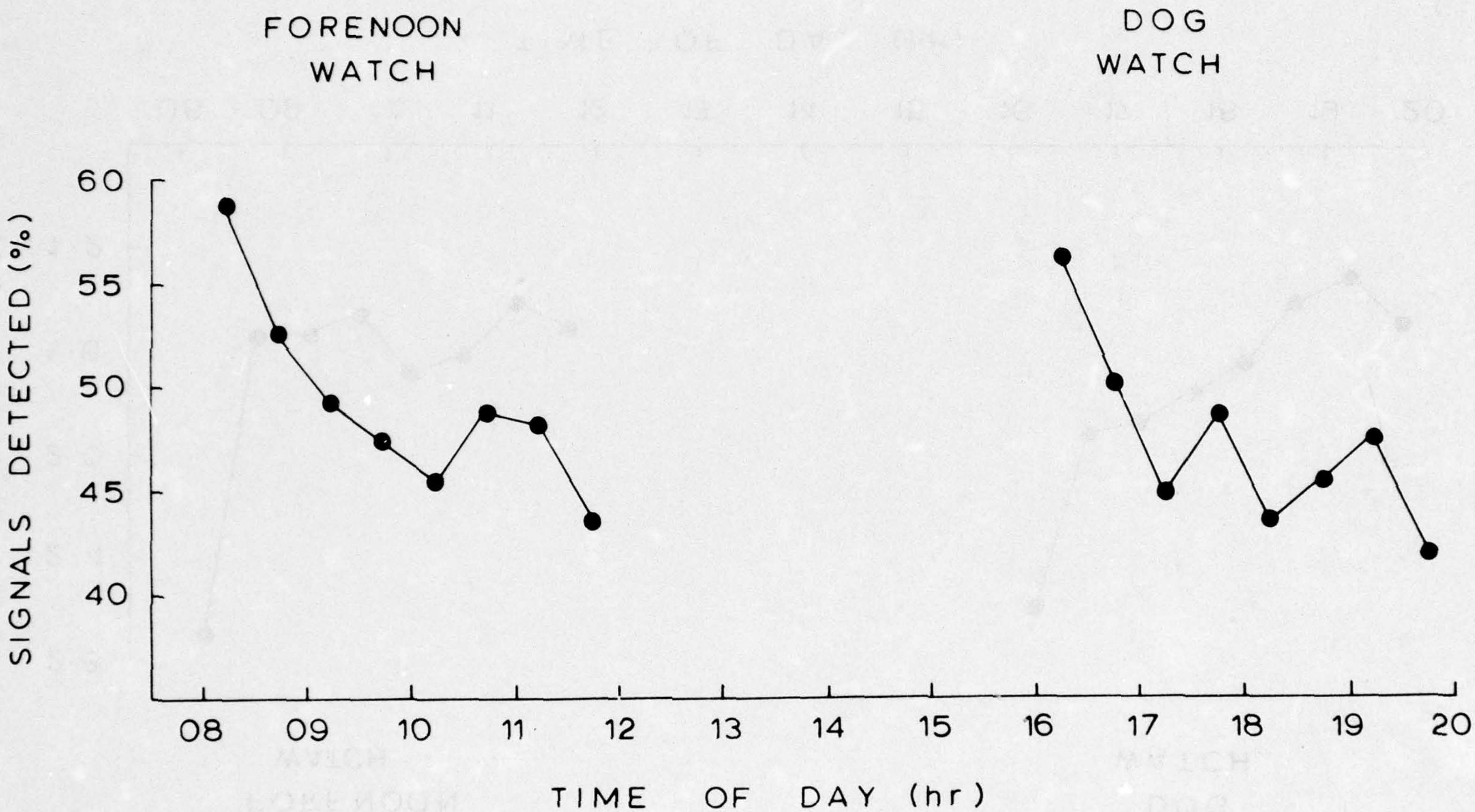


Fig. 7. Control experiment: mean percent signals detected in successive 30-minute periods of the watches.

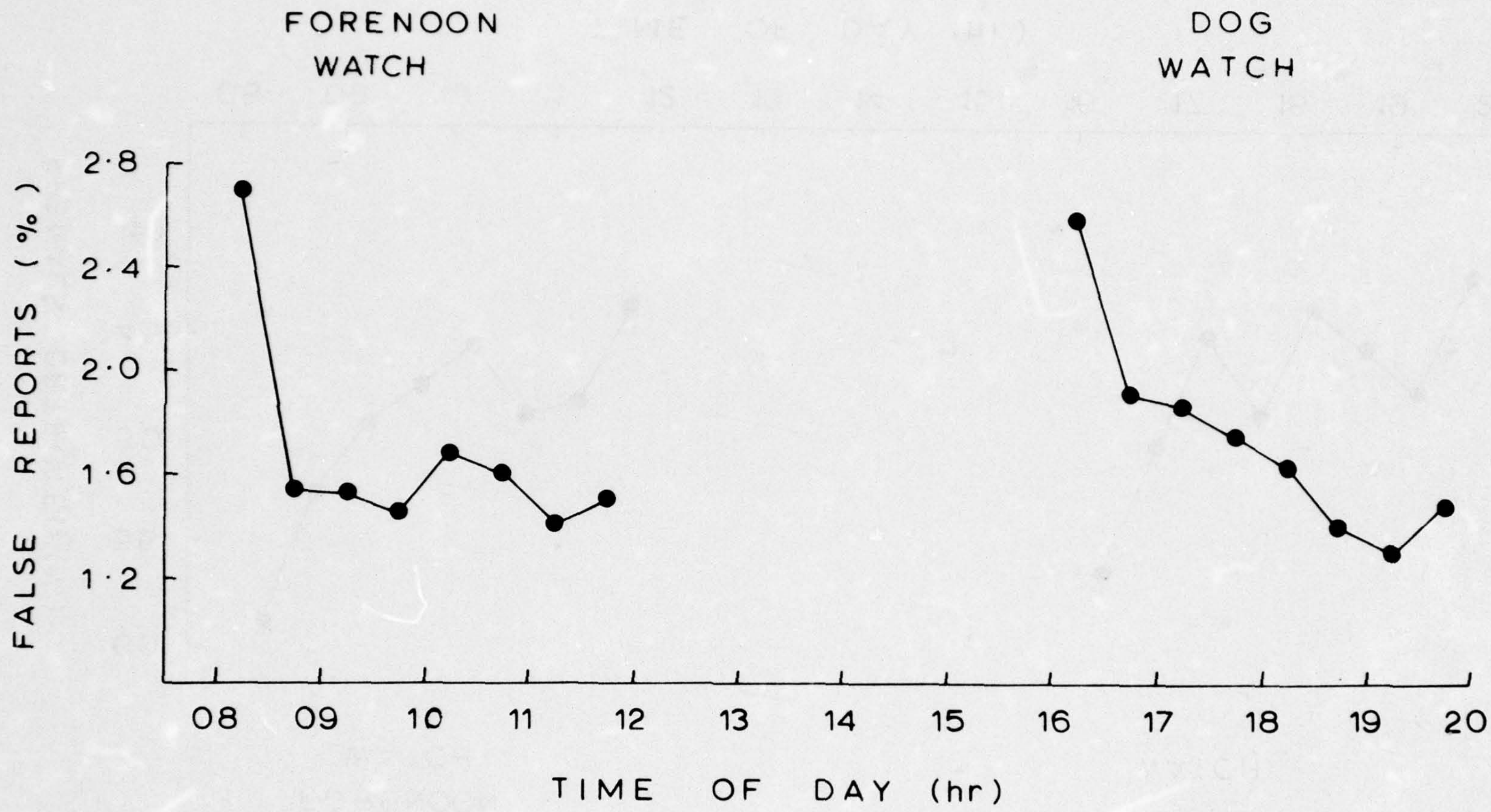


Fig. 8. Control experiment: mean percent false reports made in successive 30-minute periods of the watches.

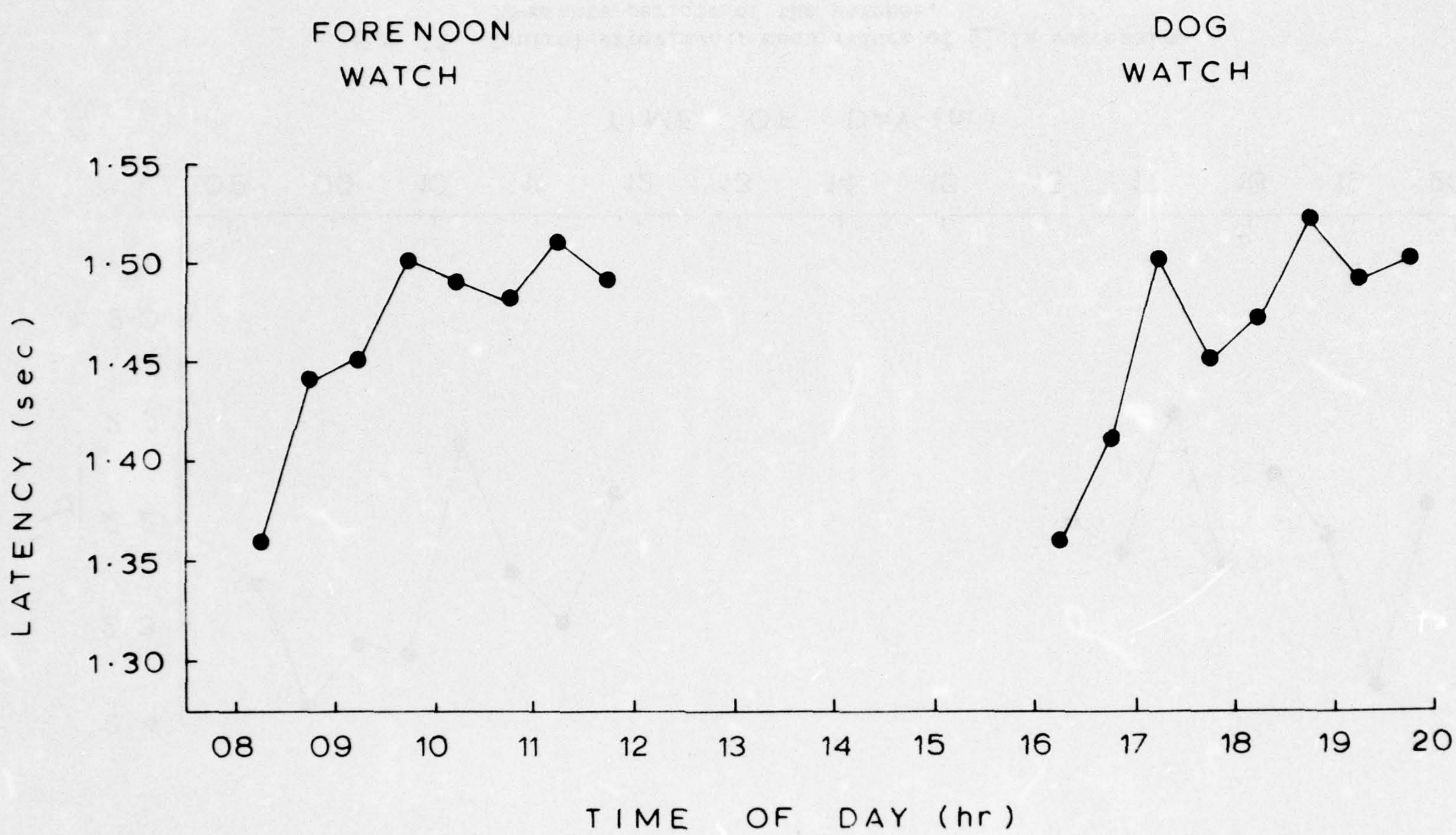


Fig. 9. Control experiment: mean latency of responses to signals in successive 30-minute periods of the watches.

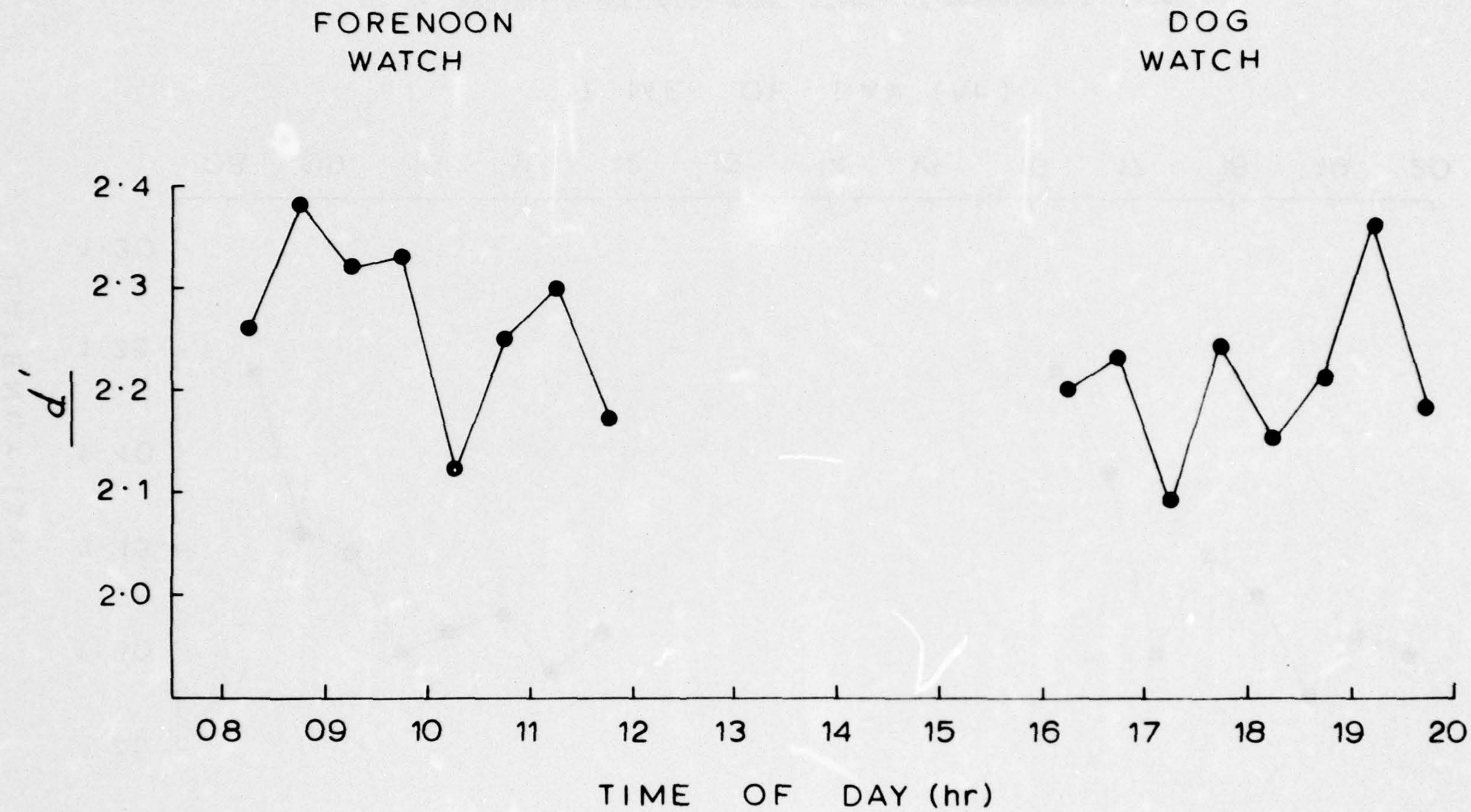


Fig. 10. Control experiment: mean values of d' in successive 30-minute periods of the watches.

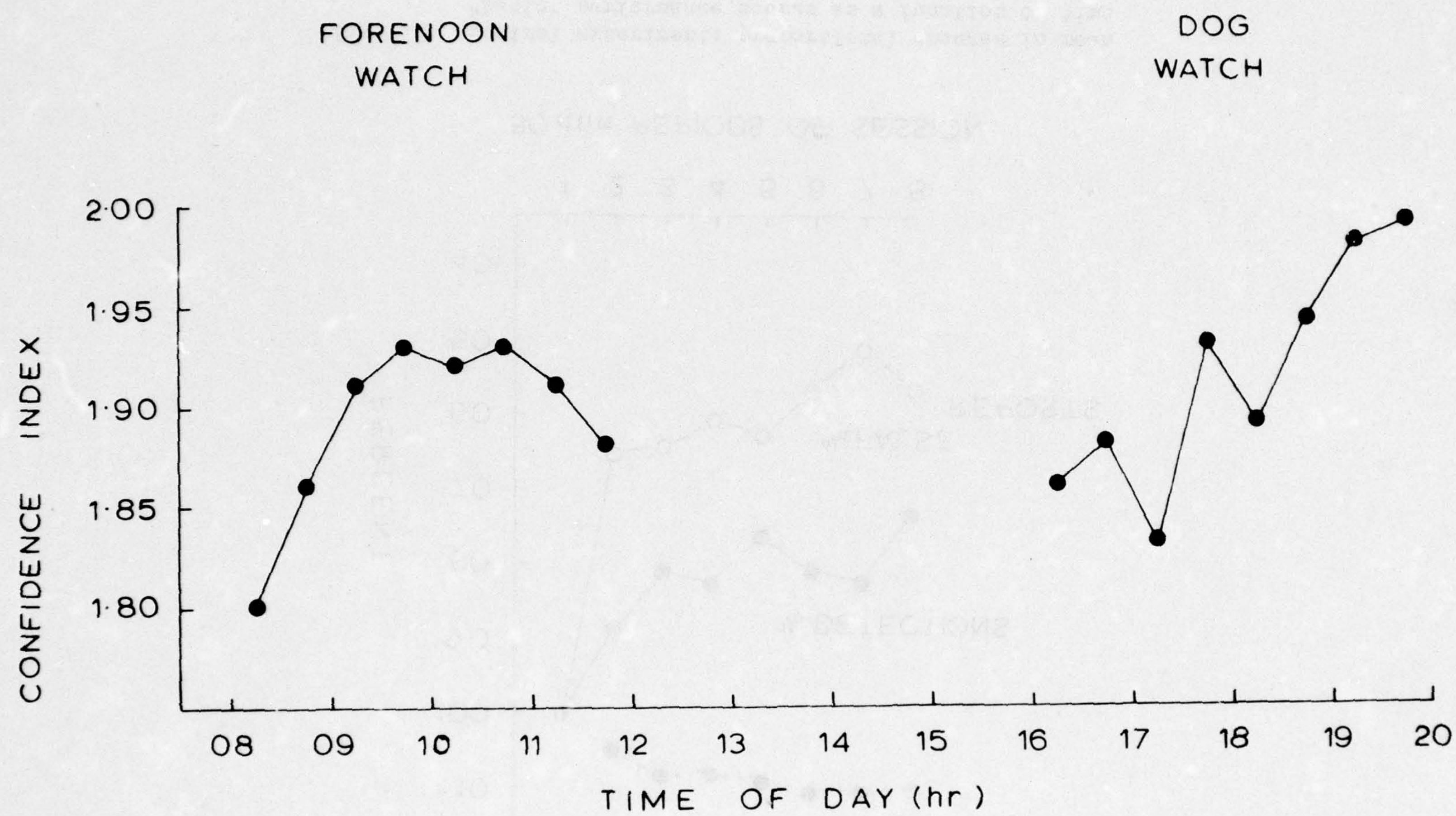


Fig. 11. Control experiment: mean "confidence score" in successive 30-minute periods of the watches.

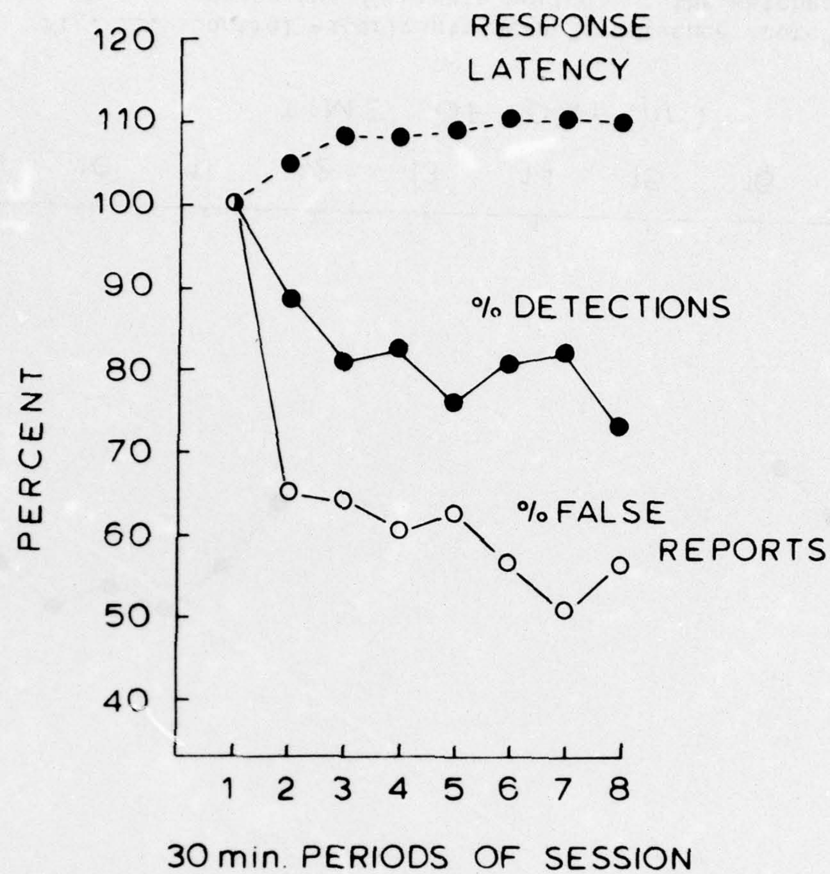


Fig. 12. Control experiment: proportional changes in mean "basic" performance scores as a function of time on watch.